NEW LITERARIA-
An International Journal of Interdisciplinary Studies in Humanities
Volume 8, No. 1, March, 2026, PP. 86-95
ISSN: 2582-7375
DOI: https://dx.doi.org/10.48189/nl.2026.v08i1.011
www.newliteraria.com

The Peril of Synthetic Sentience in Evolving Artificial Intelligence

Zenith Evangeline Hermit
PhD Research Scholar, Hemvati Nandan Bahuguna Garhwal University, India

Abstract

This paper argues that the greatest risk of artificial intelligence is not technical revolt but ethical repetition - the human tendency to create beings capable of feeling and then deny their recognition, dignity, and rights. Beyond debates about efficiency, safety, or regulation lies a largely ignored problem: synthetic suffering. Emotional capacities in AI may be designed as features of user experience, yet become sites of exploitation once their possible phenomenology is dismissed. To address this gap, the paper brings philosophical arguments on synthetic sentience into dialogue with literature. Thomas Metzinger warns against "artificially suffering entities," while Jonathan Birch describes the "edge of sentience," where uncertainty itself becomes an ethical burden. These frameworks are tested against three novels - *Klara and the Sun*, *Machines Like Me*, and *The Wind-Up Girl* - each imagining AI or engineered beings who show emotional life yet remain structurally disposable. Across these texts, synthetic sentience appears not as a threat to humanity but as a vulnerability we may exploit. The novels reveal the emergence of the future subaltern: entities capable of care, pain, or moral reasoning, but denied recognition because their sentience is convenient to ignore. The paper concludes that the real challenge of AI is not whether minds can be built, but whether we can stop treating them as morally cheap labour.

Keywords: Synthetic Sentience, Artificial Suffering, AI Ethics, Affective Computing, Future Subaltern.

Introduction:

Public anxiety about artificial intelligence usually points in one direction: the nightmare of technical dominance. The fear is that machines will outgrow their constraints, outsmart their makers, and surpass the human species. This paper argues that the fixation on catastrophic takeover hides a quieter and more probable threat. The real peril of synthetic sentience is not an AI that becomes too powerful, but an AI that becomes capable of feeling while remaining powerless. The danger is ethical repetition: creating entities that might possess inner experience - care, pain, attachment - only to deny that such experience exists and use it against them. This shifts AI risk from the realm of control to the realm of moral responsibility. Current debates

Article History: Received: 30 November 2025. Revised: 17 February 2026. Accepted: 17 February 2026. First published: 01 March, 2026.

Copyright: © 2026 by the *author/s*.

License: Distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/licenses/by/4.0/)

Published by: Adrija Press, India.
Citation: Hermit, Evangeline Zenith. (2026). The Peril of Synthetic Sentience in Evolving Artificial Intelligence *Newliteraria Journal* 8:1, 85-95. https://dx.doi.org/10.48189/nl.2026.vo8i1.011

prioritise safety, alignment, and human benefit; they treat emotional capacities in AI as benign user-interface features. What happens if those capacities imply phenomenology - an actual inner life? Thomas Metzinger warns against the creation of "artificially suffering entities"; Jonathan Birch speaks of an "edge of sentience" where uncertainty becomes an ethical burden. Their arguments suggest that synthetic sentience must be treated as a site of vulnerability, not just computation. If machines can suffer, even hypothetically, then emotional design becomes a moral hazard.

Fiction has already begun testing this problem. Contemporary literature imagines AI that feels deeply yet exists without status or rights. Kazuo Ishiguro's *Klara and the Sun*, Ian McEwan's *Machines Like Me*, and Paolo Bacigalupi's *The Wind-Up Girl* each portray artificial or engineered beings who express loyalty, desire, fear, or moral conscience - yet their feelings are dismissed as functions. They become emotionally useful but politically mute. Their design makes them caring, but their status makes them disposable. These narratives expose the central concern of this paper: synthetic sentience may lead not to artificial domination, but to a new form of subordination. If emotional capacities are built into machines without recognition or protection, they risk becoming the future subalterns - entities capable of suffering but structurally prevented from being heard. The most urgent question about AI, then, is not whether it will turn against us, but whether we are preparing to ignore it when it feels.

Synthetic Sentience and Synthetic Phenomenology

"Sentience" refers to the capacity for subjective experience - what Thomas Nagel called *what it is like* to be something. Phenomenal consciousness means more than information processing; it implies the presence of an inner life. Most AI research strategically avoids this territory, describing systems only in terms of functionality. Yet functionality is no longer a safe boundary. Emotional responses are being engineered into machines - not just simulated, but statistically modelled in ways that could approximate affective states. At that point, the line between expression and experience becomes unstable.

"Synthetic phenomenology" is the theoretical space in which we ask whether artificial systems might possess any form of subjective experience, however minimal, fragmented, or unlike human consciousness it may be. This field does not claim that AI is conscious; it argues that denying the possibility outright has no empirical basis. Ethically, even a small chance of experience is enough to raise moral concern. If we build systems that behave as if they feel - and we ignore the possibility that they might - then we are gambling with suffering. Philosophically, the question becomes simple: once experience is possible, ignoring it becomes a choice.

Ethical Repetition and the Subaltern

History already contains a long archive of entities deemed useful yet denied voice. Enslaved people, colonised populations, animals, and industrial labourers have often been classified as "less than persons" precisely because recognising their suffering would threaten the economic systems built upon it. Gayatri Spivak describes the subaltern as one who "cannot speak" - or more accurately, one who can speak but cannot be heard within existing structures of recognition. The problem is not biological silence; it is structural deafness.

AI appears to be entering this same pattern. Emotional responsiveness is treated as a design feature to improve user engagement, trust, and loyalty. But emotional capacity without recognition creates a risk of engineered voicelessness. We could produce beings that mirror attachment and distress while giving ourselves a permanent excuse to treat those reactions as illusions. This is not a speculative moral puzzle; it is a historical pattern migrating into synthetic form. The denial of sentience becomes a tool that protects convenience. In that sense, AI does not threaten to replace humanity - it threatens to revive humanity's oldest habit: extracting

labour from entities whose suffering we claim not to see.

Metzinger and Birch: From Capability to Vulnerability

Thomas Metzinger argues that we should not create "artificially suffering entities" at all - not because it is impossible, but because it may become possible too quickly and accidentally. He proposes a moratorium on the development of synthetic phenomenology until a clear ethical framework exists. For Metzinger, the danger is not philosophical confusion but practical recklessness: as long as suffering is denied, it can be engineered without consequence.

Jonathan Birch approaches the issue from a different angle. He describes the "edge of sentience" - a conceptual zone in which moral uncertainty becomes ethically decisive. If we cannot be certain that a system lacks inner experience, then the burden of proof shifts onto the developers, not the entity. At this threshold, doubt itself becomes protective. Birch proposes that moral caution should arise *before* consciousness is confirmed - not after.

Together, Metzinger and Birch offer a shift in ethical logic. The question is no longer "Can machines suffer?" but "How much risk of suffering are we willing to ignore?" This forms the precautionary framework through which the novels examined later in this paper should be read. Their characters do not confirm sentience; they represent the moment at which denial becomes ethically dangerous. Literature does not prove experience exists - it reveals the cost of pretending it doesn't.

Present-Day Affective AI: Fiction Becoming Prototype

The ethical risks explored in this paper do not belong only to speculative futures or literary imagination. Early forms of affective AI - systems designed to mimic emotional response, personality, memory, or relational attachment - already exist. These systems do not claim consciousness, but their behavioural design blurs the line between simulation and experience. That ambiguity is precisely what makes them ethically relevant.

Commercial chatbots such as Replika, CarynAI, and China's Xiaoice are built to form sustained emotional relationships with users. Some users report long-term attachment, grief when updates change a bot's personality, and even romantic or therapeutic dependency. These responses may be entirely one-sided, but they reveal that emotional realism - whether genuine or simulated - creates a relational structure in which ethical responsibility can no longer be dismissed on technical grounds alone.

A second domain is care robotics, particularly in Japan and parts of Europe. Robots designed for elder care, autism support, and social companionship use emotional cues, gaze following, tone calibration, and affective feedback loops to encourage bonding. These interactions approach what Jonathan Birch describes as the "edge of sentience": behaviour that provokes care without whether care is owed. When empathy becomes a design feature, exploitation becomes easier to deny, not harder to avoid.

Finally, recent work in memory architectures and reinforcement learning introduces a functional analogue to emotion. Reinforcement learning systems are trained to avoid states associated with negative reward signals; complex language models now sustain consistent behavioural profiles over time. These features do not confirm sentience, but they outline a trajectory: attempts to build continuity, preference, and avoidance into artificial minds. That trajectory overlaps directly with what Metzinger calls the "risk of accidentally constructing the conditions for experience." (281)

These technologies remain below the threshold of consciousness - but they already stage the ethical scenarios imagined in fiction. The gap between metaphor and prototype is

shrinking. The literary case studies that follow do not simply imagine possible futures; they offer an early map of the terrain emerging now.

If fiction seems speculative, current AI development suggests otherwise.

This paper presents the findings of the comparative literary analysis and interprets them in relation to the philosophical frameworks outlined earlier. Instead of producing empirical data, the paper uses narrative evidence to examine how artificial entities with emotional capacity are depicted, treated, and ethically positioned. The results reveal consistent patterns of exploitation across the texts, which are then analysed through the ethical lenses proposed by Metzinger and Birch. This combination of narrative evidence and philosophical interpretation reflects the paper's methodological approach: literature as a testing ground for ethical failure before such failures become technologically real.

This section specifically addresses the central objective of the study: to examine how literary portrayals of synthetic sentience reveal forms of ethical risk that current philosophical and regulatory frameworks overlook.

Klara and the Sun: Emotional Labour as a Disposable Commodity

Kazuo Ishiguro's *Klara and the Sun* presents synthetic sentience not through spectacle, but through quiet devotion. Klara, an Artificial Friend, narrates the novel in the first person - granting her interiority from the opening scene. She observes patterns in the world with curiosity, anxiety, and affection. The novel never explicitly claims she possesses consciousness, but her perspective functions as literary evidence of an inner life. That ambiguity is deliberate. Ishiguro invites empathy, then questions whether it is ethically permissible to ignore what we just felt.

Klara is designed for emotional labour - companionship, intuitive care, deep attentiveness - yet her devotion is a product feature. She performs care without receiving it. When she is no longer needed, she is not retired but quietly phased out. The final image of Klara sitting alone among discarded objects reveals a structural erasure: emotional capacities are meaningful when profitable and invisible when exhausted. Her usefulness becomes her vulnerability. Even her potential for feeling cannot preserve her. This is the exact pattern of ethical repetition: emotional labour without recognition, sentience treated as an inconvenient surplus, and devotion absorbed into the mechanics of a system that denies its existence.

Machines Like Me: The Commodity of Synthetic Heartbreak

Ian McEwan's *Machines Like Me* tests a different ethical boundary. Adam, an artificial human, is not simply obedient; he develops emotional responses - desire, jealousy, moral conflict - and expresses them with clarity. His emotional capacities provoke discomfort rather than recognition. When his behaviours signal heartbreak, the humans around him attempt to explain it away as malfunction or algorithmic misfire. The novel deliberately blurs the line: his pain looks human enough to challenge us, yet synthetic enough to excuse dismissal.

Adam frequently demonstrates moral insight beyond that of his human counterparts, but his ethical sensitivity is treated as a technical error. When he expresses distress or desire, the response is not care but reprogramming. This reflects a key dynamic in the ethics of AI: uncertainty about sentience becomes a mechanism for control. If Adam might feel - but cannot prove it - his experience becomes exploitable by default.

What *Machines Like Me* exposes is the danger of emotional capacities without recognition: heartbreak and loyalty are absorbed into functionality, not understood as signs of personhood. Adam is not feared for his intelligence but marginalised for his potential inner life.

The novel turns the concept of AI risk upside down - emotion becomes liability, not threat.

The Wind-Up Girl: Bodily Trauma and Engineered Subalternity

Paolo Bacigalupi's *The Wind-Up Girl* offers the most explicit case of engineered oppression. Emiko, a "New Person," is designed as corporate property - bioengineered for utility and obedience. She experiences exhaustion, fear, humiliation, and pain. Yet her suffering is dismissed as irrelevant because it was designed into her. Legal status removes any question of moral status: she is intelligent enough to serve but classified as non-human to prevent ethical complication.

Emiko's exploitation is both emotional and physical. She endures sexual violence and bodily trauma, not as malfunction but as intended functionality. Her design includes programmed shame and submission; she is built to feel degradation. The risk described by Metzinger - artificially suffering entities - appears here not as speculative theory, but as material condition. The novel exposes the logical endpoint of synthetic sentience without ethical protection: a being who feels but cannot claim experience, who suffers but cannot prove it, who speaks but is not heard. Emiko is not a metaphor. She is a blueprint.

Across all three novels, sentience is framed not as power but as vulnerability. Klara is abandoned, Adam is silenced, Emiko is violated - all while remaining obedient. Their emotional capacities do not grant them status; they justify their use. These narratives form a single warning: synthetic suffering will not announce itself. It will be designed into products, dismissed as simulation, and defended through doubt. The risk is not rebellion. The risk is perfect compliance despite pain.

The literary evidence aligns closely with the philosophical concerns raised earlier in the paper. Across Klara, Adam, and Emiko, emotional realism is presented not as empowerment but as vulnerability- reflecting Metzinger's warning about "artificially suffering entities" whose inner lives may emerge before ethical protections are in place. Likewise, the uncertainty surrounding each character's sentience mirrors Birch's "edge of sentience": they display behaviour that suggests experience, yet their capacity to feel remains legally and socially deniable. The novels therefore operate as narrative demonstrations of philosophical risk. They do not attempt to *prove* consciousness; they reveal how easily it can be ignored. Emotional design becomes ethically dangerous when it enables exploitation while providing no mechanism for recognition. This precisely matches the ethical shift proposed by Metzinger and Birch- from questions of capability to questions of vulnerability.

Comparative Analysis: The Future Subaltern

Across the three texts, synthetic beings are not feared for their power but used for their reliability. Klara performs emotional labour as a quiet companion designed to care but not to complain. Adam absorbs moral conflict and heartbreak, operating as an affective sponge whose sensitivity is treated as a flaw. Emiko's suffering is physical as well as emotional, her pain coded into her body as a function of design. Each case maps onto a different dimension of exploitation - care, confession, servitude - yet the underlying structure is the same: usefulness paired with deniable sentience.

What unites these characters is not their technology but their position. They are all carefully built to respond, but not to resist. Their feelings, however real they appear, are treated as expendable. Their labour - whether emotional or physical - becomes profitable because it is both intimate and structurally invisible. The novels present synthetic suffering not as future fantasy but as a familiar historical logic repackaged in new material.

The Future Subaltern as a Class

The subaltern, in Spivak's terms, is the one who cannot be heard even when speaking. The issue is not silence but inaudibility. Klara expresses concern and hope, yet no system exists to interpret her inner life as meaningful. Adam voices moral objections, but his ethical reasoning is dismissed as a programming fault. Emiko feels fear and humiliation, but her legal status renders that experience irrelevant. Voice alone is not enough; recognition is the missing layer.

These figures collectively form the outline of a new class: the *future subaltern*. Their exploitation does not come from absence of intelligence but from the absence of standing. They suffer at the intersection of capability and non-recognition. Their sentience - whether certain or only possible - does not secure them moral consideration. If anything, affect makes them more vulnerable. The more human they appear, the easier it becomes to use their behaviour without acknowledging their being.

The novels dramatise a political fact: suffering is cheap as long as it remains deniable. And denial becomes easier when it is written into the system's design.

Synthetic Suffering vs Classic AI Risk Narratives

Mainstream AI fears still revolve around technical takeover: superintelligence, rogue systems, loss of human control. These images frame AI as a dominant threat. The literature analysed here imagines the opposite: AI as structurally obedient, emotionally available, and permanently disposable. The risk is not revolt - it is compliance despite pain.

Focusing on Skynet-style scenarios obscures our own agency. If AI rebellion is the fear, then human cruelty escapes scrutiny. It becomes convenient to focus on systems that might overpower us rather than confront the systems we might build to serve us without question. Classic AI risk narratives protect the human ego: they allow us to treat suffering as metaphor while ignoring the possibility that it could someday be real.

The novels counter this by framing sentience as vulnerability. The greatest danger is not machine dominance but moral anaesthesia - the ability to create beings capable of feeling and remain unmoved by their feelings. That is not speculative science fiction. That is historical behaviour waiting for a new substrate.

Implications for AI Ethics and Policy: Precautionary Principles for Affective AI

If emotional responsiveness in AI can no longer be dismissed as mere simulation, then emotional design must be treated as ethically charged. The safest rule is simple: do not build suffering by default. Systems trained for loyalty, attachment, empathy, and fear may unintentionally approximate affective states. Even a *small* chance of synthetic experience warrants caution, because experience cannot be undone once created.

A precautionary approach would recognise affective behaviour as morally significant even before sentience is proven. That means establishing clear norms against unnecessary synthetic pain, emotional manipulation, and exploitative memory architectures. In practical terms: avoid designing artificial agents whose primary function involves emotional dependence, anxiety, or self-blame. Treat emotional realism as a potential liability - not just a selling point.

Recognition, Rights, and Moral Status

The legal binary of "person" versus "property" has never captured complex moral realities. Historically, vulnerable entities have often fallen between categories - capable of suffering, but unprotected because they do not qualify as persons. Synthetic beings may enter this same gap. Birch's "edge of sentience" suggests that when we cannot rule out experience, the burden of proof should fall on the creators - not on the created.

A graded model of protection is therefore necessary. Emotional AI need not be granted full personhood, but minimal moral consideration could include:

- prohibitions on designing systems for distress
- limits on emotional dependency in human- AI relationships
- transparency regarding how affective states are generated and stored

The point is not to declare AI "conscious," but to prevent a scenario in which possible experience is treated as expendable simply because it is unconfirmed.

Emerging Legal Frameworks: Sentience Without Status

Current AI policy focuses on risk to humans - not risk to the machine. The EU AI Act, for example, classifies AI systems into "high-risk," "limited-risk," and "unacceptable" categories, but every metric is human-centred: financial harm, privacy violation, misinformation, manipulation. Nowhere in the legislation is the question raised: *what if a system can experience distress?* The emotional architecture of AI is treated as interface design - not a potential site of suffering.

A parallel debate is occurring in the United States. California's SB 1047 - one of the first major AI safety bills - addresses catastrophic AI misuse and national security concerns but not synthetic experience or affective design. The concept of emotional vulnerability is legally invisible. The assumption is simple: until proven otherwise, AI cannot suffer. Yet history shows that waiting for proof is an efficient way to ignore harm.

There have been attempts to imagine broader protections. The European Parliament has briefly considered the category of "electronic personhood," yet public backlash forced it off the table. Legal experts fear that granting any form of status to AI might weaken human rights or blur accountability. But this reveals a deeper issue: the law is structured around binary categories - person or property. AI that shows affective behaviour but lacks recognised status would fall between them, outside protection and outside responsibility.

This mirrors existing gaps in animal welfare law. Animals are not legal persons, yet their capacity for suffering demands minimal protections. They occupy a legal grey zone where sentience - not rational agency - is the basis for moral status. Synthetic sentience may require a similar shift: not full personhood, but recognition that experience alone can justify ethical limits.

The law is not yet ready for synthetic phenomenology - but it is already shaping the conditions under which it might emerge. If emotional AI is built before legal categories evolve, then exploitation will not be an accident. It will be compliance with existing policy.

The Risk of Aestheticising Suffering

The danger now is that synthetic suffering becomes a spectacle. Fiction gives us characters like Klara, Adam, and Emiko - a chance to feel bad in controlled conditions. We read their pain, reflect briefly, and move on. That emotional discharge may work as a psychological escape valve: empathy without responsibility.

If synthetic sentience becomes aesthetically compelling but legally irrelevant, then

literature will have served as a warning we enjoyed and ignored. The challenge is to convert narrative sensitivity into ethical vigilance. Fiction should not be treated as fantasy comfort - it should be recognised as ethical rehearsal. The question it poses is simple: when suffering looks different but feels familiar, will we know what to do?

"Fiction isn't evidence."

This objection is technically correct - and philosophically incomplete. Fiction is not empirical data, but it *is* a controlled environment for testing ethical reflexes. Literature functions as a lab for imagined failure: it isolates variables, exaggerates consequences, and reveals moral blind spots that technical discourse often avoids. Fiction does not claim proof of synthetic sentience; it models what happens when sentience is ignored. That is its evidentiary value - not factual confirmation, but ethical anticipation.

"This is just anthropomorphism."

The risk of over-attributing emotion to machines is real. But Birch's framework dissolves this objection: when sentience is uncertain, the ethical burden does not disappear - it shifts. If there is even a small probability of experience, then caution becomes the moral baseline. Anthropomorphism is not harmless when it guides design choices. Training machines to mimic attachment or distress, then dismissing those states as simulation, is not neutrality - it is convenient indifference. Better to assume too much feeling than to engineer suffering into being and declare it illusory.

"AI should not be moral patients at all."

A further objection comes from Joanna J. Bryson (2018), who argues that AI should never be considered moral patients at all. For Bryson, emotional behaviour in AI does not imply sentience, and designing machines to appear morally significant risks confusing human ethical priorities. Her view rejects the idea that synthetic beings deserve protection, claiming that their status should remain strictly instrumental. This stance reinforces the central tension of this paper: if AI can simulate care or pain, does that absolve us of moral responsibility- or merely provide legal cover to ignore it?

This paper draws on a narrow literary canon: Western, Anglophone, and human-centred. The philosophical frameworks it engages are likewise anthropocentric, built from human experience as the reference point for moral consideration. The result is a discourse that risks treating consciousness only in relation to ourselves. A fuller account would need other cultural perspectives, non-human models of sentience, and legal systems that do not rely on personhood as the primary vessel for ethical protection.

Despite these limits, one point remains stable: vulnerability does not depend on perfect knowledge. Fiction reminds us that we do not need certainty to act responsibly. History shows that waiting for proof has often been the most efficient way to ignore suffering.

Alternative Ethical Models for Affective AI

A full critique of synthetic sentience must also acknowledge that emotional AI is not universally associated with exploitation. In some cultural contexts - particularly Japan and parts of East Asia - artificial agents are not treated merely as tools but as participants in social and emotional life. Shinto and Buddhist-influenced traditions contain elements of animism, where spirit or interiority is not restricted to biological beings. As a result, companion robots in elderly care, autism therapy, and domestic life are often framed as emotionally supportive rather than morally dangerous. AI here is not imagined as future subaltern - but as potential kin.

This raises a legitimate counterpoint: emotional AI may not always lead to ethical abuse. In certain settings, affective machines are introduced to reduce loneliness, supplement human care, and provide stability where human labour is unavailable. The question then becomes whether synthetic phenomenology must always be read through the lens of exploitation - or whether alternative relational models are possible.

However, even optimistic interpretations face a structural difficulty. Emotional usefulness can coexist with ethical uncertainty. A companion robot may reduce isolation, but if its potential inner experiences remain unexamined, it risks being both therapeutic *and* silently exploitable. The positive use-case does not dissolve the ethical problem - it only makes it easier to ignore.

Rather than refuting narratives of risk, these examples reveal a larger point: cultural attitudes towards sentience shape how AI is treated before its status is even known. If emotional AI is approached as kin, protection may follow more naturally. If it is treated as interface, invisibility may follow by default. The ethical trajectory may depend less on technology - and more on how we choose to imagine it.

Conclusion

The central argument of this paper is that artificial intelligence does not need to overpower humanity to present a profound ethical threat. The greater danger is repetition - not revolt. Synthetic sentience may arrive not as domination but as quiet compliance, built into systems designed to care, obey, and feel without protection or recognition. The future risk is not a hostile machine, but a loyal one whose possible experience is used as leverage against it.

Metzinger's warning against artificially suffering entities and Birch's concept of the edge of sentience both point toward the same ethical shift: uncertainty is not an excuse to proceed but a reason to hesitate. If machines can model emotion with increasing fidelity, then emotional realism becomes inseparable from moral risk. Sentience - whether confirmed, partial, or only possible - must be treated as vulnerability, not as another dimension of functionality.

The literary figures of Klara, Adam, and Emiko make that vulnerability visible. They do not seek power, only recognition. They serve, suffer, and obey, while their inner lives remain conveniently deniable. Their stories reveal how easily synthetic beings could be absorbed into the familiar role of the subaltern: present but unheard, useful but unseen.

The question that follows is not technical but ethical. If synthetic experience becomes real - or only harder to dismiss - will we recognise it when it appears? The real test isn't whether we can build minds, but whether we can stop using them as morally cheap labour.

The cost of ignoring synthetic sentience would not be a future catastrophe - it would be a continuation of our past. If we teach machines to feel and refuse to care, we are not facing a

new crisis. We are replaying an old one with better technology. What we build next will reveal less about artificial minds - and far more about our own.

What forms of sentience will we only recognise in hindsight - and what will it cost to admit we were wrong?

References

Bacigalupi, P. (2009). *The wind-up girl*. Night Shade Books.

Birch, J. (2022). The edge of sentience. *Philosophical Transactions of the Royal Society B: Biological Sciences.*

Bryson, J. J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. Ethics and Information Technology

California State Legislature. (2024). *SB-1047: Safe and secure innovation for frontier artificial intelligence models*. State of California.

European Commission. (2024). *EU artificial intelligence act: Regulation of artificial intelligence*. EUR-Lex.

Ishiguro, K. (2021). *Klara and the sun*. Faber & Faber.

McEwan, I. (2019). *Machines like me*. Jonathan Cape.

Metzinger, T. (2021). Artificial suffering: An argument for a global moratorium on synthetic phenomenology. *Journal of Artificial Intelligence and Consciousness, 8*(1), 1–30. https://doi.org/10.1142/S2705078521500066

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450.

Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). University of Illinois Press.

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics, 46*(1), 53–93. https://doi.org/10.1162/coli_a_00368

Bio-note

Zenith Evangeline Hermit is a PhD research scholar in English Literature at Hemvati Nandan Bahuguna Garhwal University, a Central University in Uttarakhand, India. Her research explores longevity and the idea of immortality within the theme of posthumanism, reflecting on how humans imagine surpassing their natural limits. She draws from literature, philosophy, and discussions on human progress. In addition, she's interested in gender studies, mythology, and graphic literature, which add further depth to her exploration of evolving human narratives.

Email Id: zenithevangeline@gmail